

J. Clin. Chem. Clin. Biochem.
Vol. 22, 1984, pp. 791–810

A New Procedure for Discriminating Between Two Patient Populations Using Multivariate Decision Limits: Application in the Detection and Exclusion of Alcoholism Based on Clinical Laboratory Findings

By E. Hansert, H. Federkiel and D. Stamm

Max-Planck-Institut für Psychiatrie, Munich

(Received December 27, 1982/March 14, 1983/August 13, 1984 (English version))

Summary: A simple, nonparametric discrimination procedure was developed and tested for use in discriminating between two populations, especially in those cases where the properties of the distributions lead to unsatisfactory results by classical discrimination procedures.

The procedure is based on transferring the conventional concept of a decision limit from one to several variables, with combinations (in set theory terms: unions and intersections) of multidimensional intervals serving as the discriminant regions (referred to here as simple discriminant regions).

The procedure was applied to two groups of patients, alcoholics and non-alcoholics. A level of efficiency 20% higher than that attainable with conventional procedures was obtained using only six clinical chemical parameters.

In addition, the procedure is more flexible than conventional procedures. It enables the solution of optimization problems for each of the three common discrimination criteria (sensitivity, specificity and efficiency) with no fundamental difficulties, and it is still possible to select at will the range(s) for the decision point and for the order of the discriminant regions.

For all of the clinical chemical parameters studied the distributions of the results for the two patient samples overlapped very markedly. If, for example, a sensitivity of 100% was specified, the best specificity attainable using the six parameters that discriminated best was only about 53%, and if a specificity of 100% was required, the best sensitivity was about 56%.

The analogy between the discrimination principle and the principle of a statistical test (the relationship between detection and exclusion of alcoholism) is also discussed.

Neues Verfahren zur Diskrimination zweier Patientenstichproben unter Verwendung multivariater Entscheidungsgrenzen:

Anwendung zur Erkennung und zum Ausschluß des Alkoholismus aufgrund klinisch-chemischer Befunde

Zusammenfassung: Für die Diskrimination von zwei Stichproben wurde allgemein und speziell für den Fall, daß klassische Diskriminationsverfahren aufgrund der Verteilungseigenschaften unbefriedigende Ergebnisse liefern, ein einfaches, nicht-parametrisches Diskriminationsverfahren entwickelt und getestet.

Es beruht auf der Übertragung des herkömmlichen Konzeptes einer Entscheidungsgrenze von einer auf mehrere Variablen, wobei als Diskriminanzbereiche Kombinationen (mengentheoretische Verknüpfungen) von mehrdimensionalen Intervallen auftreten (hier einfache Diskriminanzbereiche genannt).

Angewendet wurde das Verfahren auf je eine Stichprobe von Patienten, die als Alkoholiker bzw. von Patienten, die als Nichtalkoholiker eingestuft waren. Erreicht wurde mit dem einfachen Verfahren eine um etwa 20% höhere Effizienz als bei herkömmlichen Verfahren, und zwar auf der Basis von nur 6 klinisch-chemischen Kenngrößen.

Das Verfahren ist außerdem viel flexibler als herkömmliche Verfahren, da es ohne prinzipielle Schwierigkeiten gestattet, Optimierungsaufgaben für jedes der drei gebräuchlichen Diskriminationskriterien (Sensitivität, Spezifität und Effizienz) zu lösen, wobei noch freie Auswahl der Bereiche für die Entscheidungsgrenze und für die Ordnung des Diskriminanzbereiches möglich ist.

Bei der Anwendung zeigten sich starke Überschneidungen der gegebenen zwei Stichproben in allen zugrundegelegten Dimensionen. So können z. B. bei Vorgabe von 100% Sensitivität bzw. 100% Spezifität mit den 6 wichtigsten Kenngrößen maximal etwa 53% Spezifität bzw. 56% Sensitivität erreicht werden.

Diskutiert werden außerdem die Analogie des Diskriminationsprinzips zum Prinzip eines statistischen Tests (das Verhältnis von Ausschluß und Erkennung des Alkoholismus).

Contents

1. Introduction
 - 1.1 The problem
 - 1.2 The possible approaches
2. The experimental situation and the data
 - 2.1 The underlying probability theory
 - 2.2 The data
 - 2.2.1 Composition of the samples
 - 2.2.2 Simple statistical description of the samples
 - 2.2.3 Effect of the subsamples
 - 2.2.4 The distribution problem
3. Approaches to diagnosis based on probability theory
 - 3.1 The process of making a diagnostic decision
 - 3.2 Discriminant regions
 - 3.2.1 Decision limits
 - 3.2.2 Discriminant regions and discriminant functions
 - 3.3 Criteria for discriminant regions
 - 3.3.1 Sensitivity and specificity
 - 3.3.2 Analogy with the concept of a statistical test
 - 3.3.3 The efficiency of disjoint discriminant regions
4. Multivariate discriminant regions
 - 4.1 Parametric discriminant functions
 - 4.2 Multivariate decision limits
 - 4.3 Simple discriminant regions of the k^{th} order
5. Objectives of the discriminant analyses and the procedure followed
 - 5.1 The objectives
 - 5.2 Considerations in the selection of the multivariate decision limits
 - 5.2.1 Optimization when the decision limits have been specified for each variable
 - 5.2.2 Previously specified range for the multivariate decision point and the search for an optimum

- 5.2.3 The computer programme
- 5.2.4 Unlimited optimization
- 5.3 The selection of the basic parameters
6. Overview of the results
 - 6.1 Dependence of the results on the constituent combination
 - 6.2 Results for a combination of six parameters
 - 6.3 Sensitivity and specificity charts
7. Comments
 - 7.1 Simple discriminant regions of the k^{th} order and multidimensional quantiles
 - 7.2 Description and estimation
 - 7.3 Visualizing and interpreting statistical results
 - 7.4 The flexibility of the methodology and the exclusion principle
 - 7.5 Multidimensional decision limits for more than two categories
- Acknowledgments
- References

1. Introduction

1.1 The problem

The study by Stamm et al. (1, 2) raised the statistical question of the quality of discrimination between two populations of patients based on clinical laboratory findings. An important part of the problem was to obtain good discrimination using as few constituents as possible.

The usual procedure for discriminating between two populations on the basis of quantitative measurements, i.e. linear discriminant analysis, was soon found to provide unsatisfactory results. A more careful inspection of the data showed that one of the ma-

for preconditions for the use of this procedure was plainly not fulfilled, as will be discussed later. Although Solberg (3), among others, had indicated that he considered linear discriminant analysis to be quite robust, we felt there must be another approach that would yield better results.

1.2 The possible approaches

In such a situation there are three possible approaches:

1. transformation of the data to produce a more suitable data structure,
2. modification of the procedure so that it will reflect reality better, and
3. reconsideration of the objectives, with a view to developing a procedure as free as possible from preconditions that are of great consequence and difficult to satisfy.

For both theoretical and practical reasons we decided not to adopt either of the first two approaches, but rather to try to generalize concepts already commonly used in diagnosis and to apply these consistently. We found that the method of "simple discriminant regions" not only enabled much better discrimination than linear discriminant analysis, but it could also be applied quite flexibly. Furthermore, this approach reflects elementary properties of multivariate distributions and is thus also of some theoretical interest.

2. The Experimental Situation and the Data

2.1 The underlying probability theory

In this paper we are concerned solely with discriminating between the two populations of individuals referred to hereafter as alcoholics and non-alcoholics. As seen from the perspective of probability theory, the clinician must usually make a decision in a situation involving a combination of two random processes. One process is that an individual admitted to a given hospital or clinic will belong to a given category (here alcoholics or non-alcoholics) with a certain frequency. The other is that an individual in a given category will demonstrate certain phenomena with a certain probability, and the clinician wants to use the degree to which these phenomena are present as the basis for assigning the individual to category alcoholics (A) or non-alcoholics (NA).

These two processes can be combined into a single model, and, whether stated expressly or not, this is

the model on which most diagnostic decision-making is based. The model can be visualized as a random mechanism M consisting of two parts, where the first part produces an alcoholic with a probability of p and a non-alcoholic with a probability of $q = 1 - p$ (see fig. 1). If an alcoholic has been produced, then the second part of the random mechanism produces a random variable X_A that, with a certain probability $P_A(T^+)$, yields a value within a previously specified but otherwise initially arbitrary range T^+ of "positive test values." If, on the other hand, a non-alcoholic has been produced in the first part of the random mechanism, the second part produces a random variable X_{NA} that, with a certain probability $P_{NA}(T^-)$, has a value within a previously specified range T^- of "negative test values." (T^+ and T^- should be considered here as parameters of the model.)

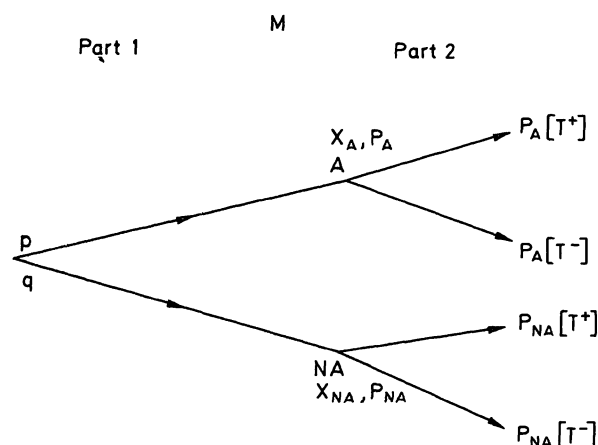


Fig. 1. Two-part, idealized probability model of patient admissions to a hospital or clinic and presence of a given illness. A = alcoholics, NA = non-alcoholics (For explanation of further symbols see text.)

If the two parts of the random mechanism are stochastically independent, the mechanism as a whole yields a value within T^+ with a probability of

$$P_M(T^+) = p \cdot P_A(T^+) + q \cdot P_{NA}(T^+) \quad (\text{Eq. 1a})$$

and a value within T^- with a probability of

$$P_M(T^-) = p \cdot P_A(T^-) + q \cdot P_{NA}(T^-). \quad (\text{Eq. 1b})$$

The probability distributions associated with X_A and X_{NA} are symbolized in figure 1 by P_A and P_{NA} , respectively.

For such a random mechanism, it makes sense to enquire into the conditional probability of being an alcoholic if the value is within T^+ , and that of being a non-alcoholic if the value is within T^- (in such a situation the usual formula $P(T^+/A)$ is justified).

These two probabilities can be obtained from the model just described, together with the definition of conditional probability, in a natural way and without resort to the controversial *Bayesian* point of view.

They can thus be expressed as:

$$P_M(A/T^+) = \frac{p \cdot P_A(T^+)}{p \cdot P_A(T^+) + q \cdot P_{NA}(T^+)} \quad (\text{Eq. 2a})$$

$$P_M(NA/T^-) = \frac{q \cdot P_{NA}(T^-)}{p \cdot P_A(T^-) + q \cdot P_{NA}(T^-)} \quad (\text{Eq. 2b})$$

Such conditional probabilities or the estimations of such values on the basis of empirical distributions are referred to in diagnosis as "predictive values" (of T^+ or T^-). Here it is generally assumed that T^+ and T^- are disjoint and include the whole range of values.

The two-part random mechanism just outlined is an idealized probability model for the situation at a hospital to which the alcoholics and non-alcoholics are admitted in a random order and are to be diagnosed as either alcoholics or non-alcoholics. The probability p is then the parameter in the model that tells how many alcoholics can be expected among the persons examined. The extent to which the assumptions hold that are made in the model, and which must hold if the use of "predictive values" is to be meaningful, cannot be assessed here. However, it is conceivable that in practice it is not unusual for the two parts of the random mechanism to be stochastically dependent.

The present paper is concerned mainly with the second part of the random mechanism. Thus the objective was not to determine the frequency of alcoholics in specific hospitals or to obtain any other information related to this frequency but rather, on the basis of carefully selected groups of alcoholics and non-alcoholics, to study statistical properties of certain parameters, i.e. of random variables X_A and X_{NA} (in

this connection statisticians sometimes speak of analysing "initial samples").

2.2 The data

2.2.1 Composition of the samples

The procedure used to select the two types of patients and the clinical chemical parameters studied have been described elsewhere (1). Quantitative determinations served as the basis for most of the laboratory findings, many of them being determinations of concentrations with the typical scale that can be regarded as quasicontinuous and as having a natural zero and, from the measurement perspective, an associated detection limit. (The complete characterization of such scales in terms of measurement theory – see e.g. *Suppes & Zinnes* (4) – will not be attempted here.)

The samples of alcoholics and non-alcoholics included patients from three and two clinical institutions respectively. Table 1 shows the numerical composition of the two groups.

Tab. 1. Patients studied (exclusively men).

Hospital	Alcoholics	Non-alcoholics
Munich-Schwabing City Hospital	27	30
Haar Regional Hospital	20	40
Annabrunn Alcoholism Treatment Centre	35	—
Σ	82	70

2.2.2 Simple statistical description of the samples

Table 2a shows several of the clinical chemical parameters studied – those that were found to be of greatest importance in discriminating between alcoholics and non-alcoholics – and the values for a few basic variables indicating the location of each distribution on the measurement scale and the amount of

Tab. 2a. A few statistical parameters describing the empirical distributions of those clinical laboratory findings that enable the best discrimination between alcoholics (A) and non-alcoholics (NA).

Statistical parameters	γ -Glutamyl-transferase (U/l)		Aspartate amino-transferase (U/l)		Alanine aminotransferase (U/l)		MCV (fl)		Creatinine (μ mol/l)		Urea-N (mmol/l)	
	A	NA	A	NA	A	NA	A	NA	A	NA	A	NA
Smallest value	11	8	7	6	7	4	73	72	39.8	42.4	0.7	1.8
First quartile	30	13	13	8	14	10	95	87	61.0	68.9	2.5	3.9
Median	61	20	17	11.5	19.5	14	100	93	68.1	78.7	3.6	5.0
Third quartile	160	32	34	15	37	24	105	100	76.0	85.7	4.6	6.4
Largest value	1430	218	130	28	93	54	117	114	104.3	167.6	8.9	24.6

variation. Because the distributions were sometimes quite asymmetrical, the median was used as the location parameter rather than the mean, and the first and third quartiles for describing variation instead of the standard deviation.

2.2.3 Effect of the subsamples

When interpreting the statistical results it is, of course, important to know whether the fact that the samples of alcoholics and non-alcoholics consisted of several subsamples played a major role. This question was therefore investigated for the most important constituents. As expected, it was found that the composition of the samples did affect the quality of the discrimination, but not in a critical way. Table 2b illustrates, with the results for γ -glutamyltransferase and urea, that the level of discrimination between the combined samples has a middle position, with the discrimination for individual institutions being either somewhat better or somewhat worse.

One of the findings in this study, illustrated in table 3, is that for some of the most important constituents (with reference to discrimination) the results for different hospitals may have different locations on the scale. The levels of significance given were obtained with the *Kruskal-Wallis* test and the *Wilcoxon-Mann-Whitney* test (U-test). In all cases the subsample from the Munich-Schwabing City Hospital had larger values than the subsamples from the other two hospitals.

2.2.4 The distribution problem

As already emphasized in the introduction, it was not possible to obtain satisfactory discrimination with the usual procedures for discriminant analysis. The explanation was thought to lie in the distribution of the findings for the alcoholics and non-alcoholics

(we shall return to this topic later). Figure 2 shows the two-dimensional distribution of γ -glutamyltransferase (abscissa) and aspartate aminotransferase (ordinate) findings for the group of alcoholics. Because of the scale used, those subjects with a γ -glutamyltransferase finding above 260 U/l and/or an aspartate aminotransferase finding above 90 U/l had to be omitted. It is clear that the distribution deviates markedly from a two-dimensional normal distribution. Here again we asked ourselves whether this and other marked deviations could be due mainly to the fact that the sample consists of various subsamples, i.e. whether within the individual institutions the deviations from a normal distribution are smaller.

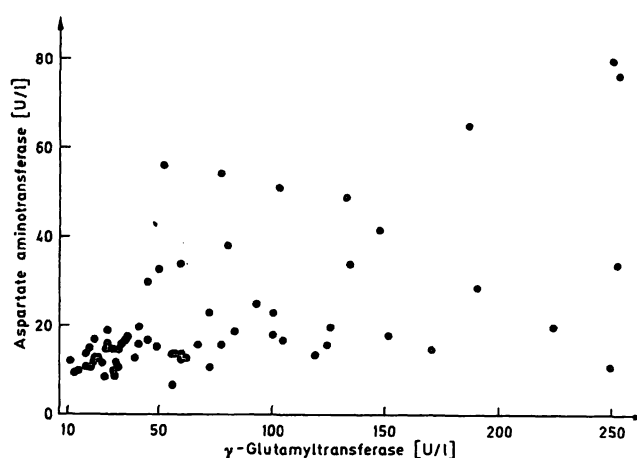


Fig. 2. Graph of the distribution of the constituent pair γ -glutamyltransferase and aspartate aminotransferase for the alcoholics studied (extreme values omitted).

A check was made of whether the results for the most important constituents were normally distributed, this being done for each subsample of subjects from the participating hospitals. Unfortunately there is no test of multivariate normal distribution available, and thus all of the checks had to be performed in one dimension. Nevertheless, if it were possible to

Tab. 2b. Medians for the subsamples for γ -glutamyltransferase (GGT) and urea-N.

Constituent		Hospitals			Combined samples
		MS	HA	AB	
GGT (U/l)	A	114	58	52	61
	NA	22	18	—	20
Urea-N (mmol/l)	A	3.2	3.9	3.6	3.6
	NA	5.9	4.6		5.0

Note: MS = Munich-Schwabing City Hospital
HA = Haar Regional Hospital
AB = Annabrunn Alcoholism Treatment Centre

Tab. 3. List of those important constituents for which the difference between hospitals was significant at the 5% level.

A = alcoholics
NA = non-alcoholics

Constituent	Category	Critical level
γ -Glutamyltransferase	A	1%
Aspartate aminotransferase	A	0.5%
Alanine aminotransferase	A	0.5%
Urea-N	NA	2.5%

show that there was a multidimensional normal distribution, it would follow that the associated one-dimensional distributions were also normally distributed. However, the check on the individual one-dimensional distributions indicated that the results for none of the subsamples, neither for the alcoholics nor for the non-alcoholics, had a multidimensional normal distribution. The data were assessed for normal distribution with a modification of the test developed by *Sarkadi & Stoermer* (see l.c. (5)).

3. Approaches to Diagnosis Based on Probability Theory

3.1 The process of making a diagnostic decision

Every diagnostic process in which a conclusion of some kind is drawn, however tentative, ends in a statement that the individual studied has a certain place in a system of categories (6). In this paper we say an individual "belongs to a particular category," and at the same time we are aware of the problems associated with this form of expression. The diagnostic process consists of a number of different elements. One of these elements, and an especially important one, is the use of already extant observations on a certain set of characteristics or measurements, for instance determinations of the concentration of certain constituents in serum. In our view, models based on probability theory are valuable here because they enable us to formalize and thus objectify the use of such observations. However, the diagnostic procedure is always carried out at a particular point in time and with a limited amount of information available: The conclusions reached may change with additional information, and the system of categories — or sometimes even the diagnostic procedure itself — may also change.

In the case under discussion, with the categories alcoholics and non-alcoholics, the system of categories consists of two elements. In principle it could be refined into more elements, but this would be predicated on a much larger data base.

A diagnostic process, however, must not necessarily lead to a "final" conclusion. Rather, categorization can be postponed and additional investigations conducted or the results of a trial treatment approach awaited. This would mean that a "grey zone" would have to be defined in the data on which the categorization process is based. Furthermore, it must be kept in mind that the categorization procedure could, in principle, include weighting factors, i.e. the data on which the categorization is based could be weighted.

It is difficult enough to achieve a good discrimination between alcoholics and non-alcoholics of even the most common and terminologically simplest kind, even when it takes an idealized form, i.e. where each observation of a certain kind made about an individual can always be assigned to a category and to one category only on a statistical basis. Therefore we will not explicitly pursue the other two possibilities.

3.2 Discriminant regions

3.2.1 Decision limits

When the system of categories consists of two categories only and a single characteristic is observed, the concept underlying the simplest possible statistical assessment is the partitioning of all possible values of the characteristic into two disjoint sets. One subset consists of the "positive" test results, a result in this subset causing an individual to be assigned to the category of persons with pathological findings (e.g. alcoholics). We shall refer to such subsets hereafter as T^+ . The other subset, consisting of the "negative" test results (e.g. non-alcoholics), includes all possible values (see Section 3.1) that do not belong to T^+ . We shall refer to such subsets hereafter as T^- . If a given characteristic is to permit differentiation between the two categories, the probability distributions associated with the categories — referred to in Section 2.1 as P_A and P_{NA} — must differ sufficiently, and the difference must sufficiently and in such a way that the difference is reflected clearly in the associated empirical distributions.

Theoretically such a difference can take numerous forms. However, for random variables such as the values for concentrations, the most important kind of difference between two distributions both theoretically and practically is that the subjects in one category (here the alcoholics) either "tend to have larger values" or "tend to have smaller values" than the subjects in the other category (here the non-alcoholics). A simple way of representing this "tendency" in probability terms is to state that for any possible value x of the random variables X_A and X_{NA} (see Section 2.1) the following holds:

$$\text{Either for every } x: P_A(X_A > x) > P_{NA}(X_{NA} > x) \quad (\text{Eq. 3a})$$

$$\text{or for every } x: P_A(X_A < x) > P_{NA}(X_{NA} < x). \quad (\text{Eq. 3b})$$

In the first case X_A is called "stochastically larger," in the second case "stochastically smaller" than X_{NA} (within the parentheses we could also write \geq or \leq ; only the strict case is of interest for the inequality

between the two sides). In such cases, a suitable value d for the characteristic is specified so that values above (or below) d are considered to be positive and all other values are negative. Such a value d is called a decision limit. Whether a measurement that, within the framework of the accuracy of measurement, falls exactly on the decision limit is regarded as positive or negative is of no theoretical interest. In the evaluations that follow such results were always classified as "positive."

3.2.2 Discriminant regions and discriminant functions

As already pointed out, in discriminating between two categories such as alcoholics vs. non-alcoholics, the ideal case is based on the assumption of a dichotomy of all possible values (all measurements) for the characteristic. This holds not only for one but also for several characteristics, i.e. also in the multivariate or multidimensional case. In order to have a general and brief way of speaking, we use the term "discriminant regions" for all such regions that serve as the basis for a discrimination rule. For continuous variables and for dichotomies, the two discriminant regions will be contiguous, i.e. they will have a "border" in common. The mathematical representation of such a "border" is called a "discriminant function." In the case of a single variable, the decision limit can be regarded as a single-valued discriminant function. Figure 3 shows an example of a discriminant function in two dimensions in which the discriminant function plays the role of the decision limit in the sense that two sides are defined, with the positive test results lying on one side and the negative test results lying on the other (the type of discriminant function used in the example was not chosen to provide any information). The discriminant regions with this property are also of great importance in multivariate cases, and one can envisage an adaptation of the inequalities 3a and 3b to such cases.

3.3 Criteria for discriminant regions

3.3.1 Sensitivity and specificity

In a categorization system consisting of two elements such as alcoholics and non-alcoholics one can expect that the two discriminant regions of positive and negative test results (T^+ and T^-) will have the following properties:

1. T^+ includes those values that can be found for alcoholics but not for non-alcoholics, and T^- includes those values that can be found for non-alcoholics but not for alcoholics.

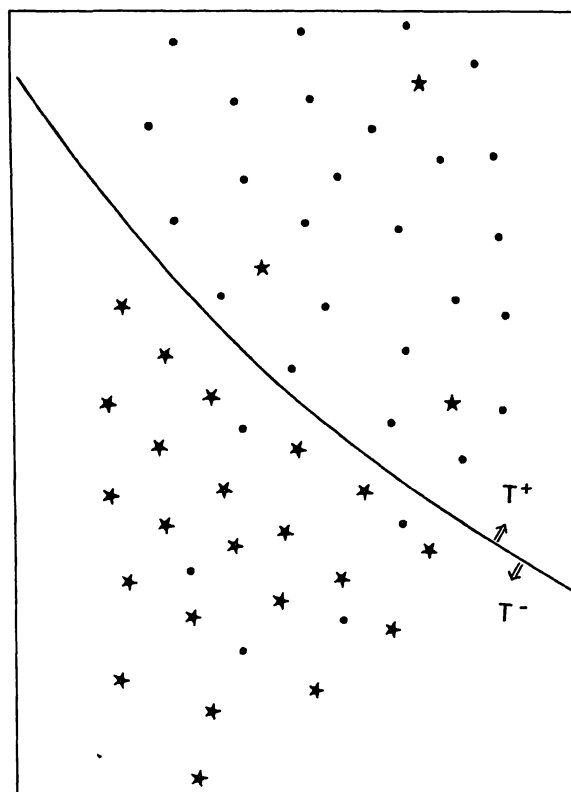


Fig. 3. Nonlinear discriminant function in two dimensions and the associated discriminant sectors T^+ and T^- .

2. T^+ and T^- include the "typical" values for alcoholics and non-alcoholics, respectively, i.e. a certain range around a location parameter for the results found in each group.

It is not immediately obvious how these properties are generally compatible with each other. A range of values that occurs only in alcoholics clearly indicates alcoholics, but it may be rare in that category, in other words, atypical. A range of values that is typical for alcoholics is not necessarily rare for non-alcoholics and vice versa. Thus two measures are required to evaluate a discriminant region, the probability that it will include the alcoholics and the probability that it will include the non-alcoholics (P_A and P_{NA}). The common terms for these measures are the "sensitivity" and the "specificity" of the range T^+ of positive test results:

$$\text{sensitivity } (T^+) = P_A(T^+) \quad (\text{Eq. 4a})$$

$$\text{specificity } (T^+) = P_{NA}(T^-). \quad (\text{Eq. 4b})$$

Symmetry can be created between the two discriminant regions T^+ and T^- if one considers that the sensitivity of T^- is identical to the specificity of T^+ and the specificity of T^- is identical to the sensitivity of T^+ .

The second of the above-mentioned requirements means that T^+ should have both high sensitivity and high specificity. The first requirement, on the other hand, is that T^+ should include all subsets with almost complete specificity and T^- should include all subsets whose complement has almost complete sensitivity. It may be advantageous to give precedence to a requirement of the second type, i.e. to choose T^+ in such a way that for a certain high specificity the sensitivity is as high as possible or for a certain high sensitivity the specificity is as high as possible. The requirement of the first type will be discussed further in Section 3.3.3.

The differences between the distributions P_A and P_{NA} can also be made use of in another way (see, e.g. (7)). In the following, however, we will discuss the diagnostic criteria sensitivity, specificity and efficiency only.

3.3.2 *Analogy with the concept of a statistical test*

A procedure for assigning things within a system of categories can also be regarded as a decision-making procedure. This suggests a comparison of the use of discriminant regions with the procedure in statistical tests (for more on the concept of a statistical test see, e.g. (8)). As a matter of fact, the analogy is complete if the examination of a particular subject is regarded as a "random experiment" in which a decision is to be made on the basis of the result(s) as to whether one of the two possible probability distributions (P_A or P_{NA}) is to be rejected or not (where non-rejection does not necessarily mean acceptance!).

To review, a statistical test consists of the following elements:

- (Null) hypothesis H_0 and alternative hypothesis H_1
- The test level: low probability α
- The critical region, causing rejection of H_0 , and, logically, acceptance of H_1 .

One can now

1. identify H_0 with P_A . Here the goal is to find a basis for deciding when to reject a combination of values as indicative of alcoholics.

Consequently: T^+ must be sensitive enough, i.e. must include enough alcoholics.

2. identify H_0 with P_{NA} . Here the goal is to find a basis for deciding when to reject a combination of values as indicative of non-alcoholics.

Consequently: T^+ must be specific enough, i.e. T^- must include enough non-alcoholics.

We have the impression that although there have been many publications on discrimination, including many statements about sensitivity and specificity, the logic behind the use of such terms has usually been understood very poorly. For this reason special attention has been paid in this paper to a thorough discussion of the analogy with the logic of statistical test procedures (see tabs. 4a und 4b). The existence of such an analogy has already been pointed out by other authors (see, e.g. (9)).

We want to emphasize once again that the non-rejection of the null hypothesis by no means has as a logical consequence the acceptance of this hypothesis. Before making and applying a discrimination rule the clinician will have to specify what the "null hypothesis" is to be. The decisive factor is exclusion. Does the clinician want to ensure that the excluded individuals are categorized as alcoholics, or as non-alcoholics? The criterion cannot be the same for both – except when both of the "test levels" (and thus also the level of efficiency) are sufficiently high. This situation is discussed in detail in the next section.

3.3.3 *The efficiency of disjoint discriminant regions*

We have seen that to evaluate a discriminant region both the probability given the category alcoholics and the probability given the category non-alcoholics are required. If two discriminant regions T^+ and T^- are disjoint, then two of the four probabilities, e.g. the sensitivity and specificity of T^+ , are sufficient. The wide use of classical discrimination procedures has led to the custom of selecting the "number of correct assignments" as the criterion for assessing the quality of the discrimination. Expressed as an equation this is:

$$\text{No. of correct assignments } C = \frac{(\text{No. of A in } T^+) + (\text{No. of NA in } T^-)}{(\text{No. of A}) + (\text{No. of NA})}$$

This concept has been defined in the same way by *Galen & Gambino* (10), without it making any basic difference whether C is given as a percentage or not. We have been unable to find a careful analysis of the theoretical properties of C , and therefore we include here at least a brief discussion of the problem of the expected value of C . A distinction must be made between the expected value in the random mechanism M , described earlier in Section 2.1, and the expected value of C in a situation such as the one discussed here, in which there is already a sample each of the populations of alcoholics and non-alcoholics of size n_A and n_{NA} , respectively. For M , i.e. for a series of

Tab. 4a. The conceptual relationship between high sensitivity and a statistical test for assigning an observed value to one of two given probability distributions.

	T ⁺	T ⁻
Given A	Highly probable High sensitivity $H_0 \longleftrightarrow P_A$	Highly improbable Critical region for P_A Test level \longleftrightarrow 100% minus sensitivity (%) Power of test \longleftrightarrow specificity (%)
	Highly probable that A will be included	Diagnosis of A is to be rejected

Tab. 4b. The conceptual relationship between high specificity and a statistical test for assigning an observed value to one of two given probability distributions.

	T ⁺	T ⁻
Given NA	Highly improbable High specificity Critical region for P_{NA} Test level \longleftrightarrow 100% minus specificity (%) Power of test \longleftrightarrow sensitivity (%)	Highly probable $H_0 \longleftrightarrow P_{NA}$ Highly probable that NA will be included
	Diagnosis of NA is to be rejected	

stochastically independent repetitions of M , the following expected value results:

$$E_{MC} = p \cdot P_A(T^+) + q \cdot P_{NA}(T^-) \\ = p \cdot \text{sensitivity}(T^+) + q \cdot \text{specificity}(T^+).$$

In the second part of the random mechanism M , i.e. in the present situation, the following expected value results:

$$EC = \frac{n_A}{n_A + n_{NA}} \cdot P_A(T^+) + \frac{n_{NA}}{n_A + n_{NA}} \cdot P_{NA}(T^-).$$

This expected value is dependent on the sample size in a way that is unnecessary. It thus appears advisable in this part to use the following measure as the theoretical criterion for the quality of a discrimination rule:

$$\text{Efficiency}(T^+) = P_A(T^+) + P_{NA}(T^-),$$

as also suggested by Büttner (11). An unbiased estimate of this theoretical efficiency in the given situation is:

$$\text{Empirical efficiency}(T^+) =$$

$$\frac{\text{No. of A in } T^+}{\text{No. of A}} + \frac{\text{No. of NA in } T^-}{\text{No. of NA}}.$$

If the samples have the same size, then the value for the empirical efficiency is identical to the value for C

except for a factor of 2. Here, again, it makes no basic difference whether the probabilities, and thus also the estimations of the probabilities, are expressed as percentages or not.

If the value for the theoretical efficiency is very high, the probability of a correct assignment is very high, regardless of whether an alcoholic or a non-alcoholic is concerned. Thus if a discriminant region with high enough efficiency can be found, then it can be used in the sense of the analogy to a statistical test both to exclude alcoholics and to include alcoholics (exclude non-alcoholics and include non-alcoholics). This means that the only basis for a symmetrical diagnostic criterion is a high enough level of efficiency in the sense described above.

4. Multivariate Discriminant Regions

4.1 Parametric discriminant functions

As emphasized in Section 3.3.2, the analogy between a two-category discrimination problem and a statistical test in its classical form is complete. Each of the two hypotheses to be tested against each other consists of a probability distribution, namely P_A and P_{NA} , respectively (testing a simple hypothesis against

a simple hypothesis). If these two probability distributions were stated explicitly, then (at least in principle) optimal discriminant regions could be calculated, and thus also optimal discriminant functions. If we have two normal distributions with the same matrices of covariance, the optimal discriminant functions are, of course, linear functions – see, e.g. (12). This therefore indicates that linear discriminant functions may be appropriate even when normal distributions are assumed but must be estimated. (If the two matrices of covariance must be regarded as different, it is advisable for the same theoretical reasons to use quadratic discriminant functions instead.) The mathematical simplicity of linear discriminant functions has led to linear discriminant analyses being included in almost all statistical programme packages, even to the extent that linear discriminant analysis is often seen as *the* multivariate procedure for discrimination.

The general applicability of a procedure makes different analyses (apparently) comparable, but it does not necessarily produce an awareness of when the procedure can be expected to be advantageous and when not. Available algorithms for linear discriminant analysis have two shortcomings. They are inflexible insofar as they permit optimization of only a single parameter, namely the number of correct assignments C (see Section 3.3.3). In addition, in spite of its mathematical simplicity a linear discriminant function remains a theoretical construct, and its relationship to the given data is usually not obvious. We will come back to this problem again – see Comment 7.4.

The quality of the attainable discrimination depends on two different factors. One is the form of the distributions for the alcoholics and non-alcoholics, respectively. The other is the distance between these two distributions or the extent of overlapping. If for a certain set of variables the values found in the samples studied do not overlap (or the overlapping of the probability distributions is slight enough to be disregarded), then credit may be given to the use of a discriminant function if the discrimination is good, even though the discrimination could actually have been carried out with the naked eye (see, e.g. fig. 3, where the discriminant function is, in fact, the result of visual inspection of the data). However, if the overlapping of the two distributions is too great to be disregarded and the form of the theoretical distribution is unknown but clearly not that of a normal distribution, then an attempt must be made to find a means of constructing discriminant regions that is as free as possible of preconditions and yet easy to carry out.

4.2 Multivariate decision limits

As just discussed, optimal discrimination can be expected with those discriminant functions in common use only if very restrictive preconditions are fulfilled. On the other hand, even linear discriminant functions are not easy to interpret. 'They represent the result of a calculation procedure into which the user usually does not have sufficient insight. This would be true also for other parametric discriminant functions (and also for some suggested nonparametric procedures, see, e.g. (13)). Therefore, even if an adequate multivariate distribution type were known, the question would have to be raised of whether for continuous measurements there are not other, directly interpretable methods for discrimination that are free from difficult-to-verify and all too restrictive prerequisites.

We set ourselves the task of transferring the very simple and easily understandable concept of the decision limit to any arbitrary number of dimensions and then of seeing what could be achieved with such a concept.

Let us assume a certain set of different kinds of observed variables. Then in a first step we can envisage a preassigned decision limit d_i for each of these variables, where values above the decision limit represent either positive or negative test results. If m is the number of variables, then after the first step we have:

$$d = (d_1, d_2, \dots, d_m) \\ T_1^+, T_2^+, \dots, T_m^+ \\ T_1^-, T_2^-, \dots, T_m^-$$

The measurements obtained for a given subject for these m variables form a point in a coordinate system

$$x = (x_1, x_2, \dots, x_m).$$

For every point x and a given d there are m statements $D_1^*, D_2^*, \dots, D_m^*$ about whether a given coordinate is "positive" or "negative":

$$x \rightarrow (D_1^*, D_2^*, \dots, D_m^* / d_1, d_2, \dots, d_m).$$

Here either a "+" or a "-" is inserted for *, depending on whether x_i is "positive" or "negative" with regard to d_i . From the statements $D_1^*, D_2^*, \dots, D_m^*$ different multivariate regions can be defined via logical connectives (see, e.g. Büttner (9)) with the following properties:

1. They are defined by a single m -dimensional point.
2. They are the result of a simple logical connective.
3. They represent connectives of multidimensional intervals such as have long been used in analytical geometry as basic mathematical elements.

In two dimensions, for example, the following connective rules are possible:

- a) $T^+ \longleftrightarrow D_1^+ \text{ and } D_2^+$
 $T^- \longleftrightarrow D_1^- \text{ or } D_2^-$
 b) $T^+ \longleftrightarrow D_1^+ \text{ or } D_2^+$
 $T^- \longleftrightarrow D_1^- \text{ and } D_2^-$.

Figures 4a and 4b illustrate these two possibilities. If there are more than two variables more compound statements are of course possible. Nevertheless, in such a case the attempt can and should be made to get along with a smaller number of variables.

In the following we will refer to all such discriminant regions that have a certain similarity to an interval and result from individual decision limits via logical connectives, i.e. that are defined by a single decision point d and the connective rules, as *simple discriminant regions*.

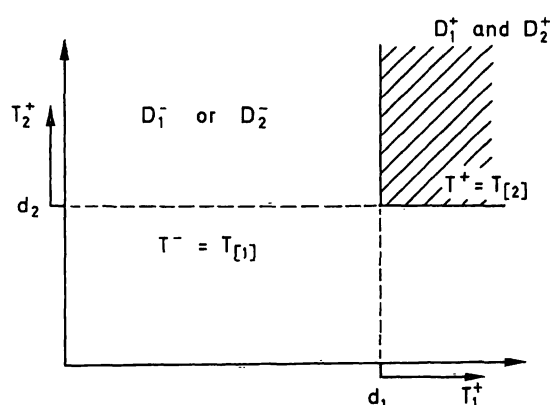


Fig. 4a. Simple discriminant sector of the second order (at least two positive findings) in two dimensions.

The symmetry of an analysis is evidenced by the fact that the results do not change when the order of the variables is changed at will. In order to be able to write down general formulations, the logical "and" connection is symbolized by " \wedge ," the logical "or" connection by " \vee " and a method of notation used that is taken from formal logic. For every variable there are two possible statements D^+ and D^- . A given connection of all m variables can be expressed, for example, as

$$D_1^+ \wedge D_2^+ \wedge \dots \wedge D_k^+ \wedge D_{k+1}^- \wedge \dots \wedge D_m^-,$$

where k should be any natural number $1 \leq k \leq m$. If the subscripts in this expression are interchanged in all possible ways, one obtains a certain set of statements with the same structure. If all of these statements are connected with " \vee ," then one gets an overall statement that is symmetrical with reference

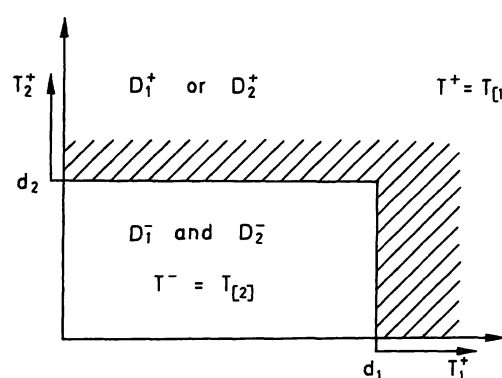


Fig. 4b. Simple discriminant sector of the first order (at least one positive finding) in two dimensions.

4.3 Simple discriminant regions of the k^{th} order

The number of logical connections with m variables (decision limits) can be reduced markedly if agreement can be reached that all of the variables are used symmetrically in the analysis. This is certainly not always justified. But on the other hand, different amounts of prior information usually lead to different views about which variables should have preference over which others. We therefore consider the specification that all variables are to be treated symmetrically as an objective specification at the very least, and one which can be rescinded if the results indicate this to be necessary, whereby the whole methodology must prove itself to be flexible enough.

to the variables. The associated simple discriminant region would consist of points x that are "positive" if exactly k coordinates exceed the associated decision limits in a "positive" direction. This would, of course, be too restrictive for a T^+ region. Rather, the procedure just described would be carried out for all values of k that exceed a certain minimum value and the resulting statement sets would then be connected again with " \vee ." An explicit formal representation of the resulting general expression will not be given here, however. The procedure described amounts to defining a T^+ region by specifying that a point x counts as positive if *at least* k coordinates exceed the associated decision limits of the decision point d in a positive direction. We refer to a discriminant region established in this manner as a "simple discriminant

region of the k^{th} order" and symbolize it by $T_{[k]}$. If $m = 2$, for the first example in Section 4.2 T^+ is a $T_{[2]}$, and for the second example T^+ is a $T_{[1]}$ — see figures 4a and 4b. It is clear that all discriminant regions of this type are symmetrical with reference to the variables. For m variables there are therefore just as many different ways of using a simple discriminant region that is symmetrical in this regard. The final point to point to be made is that T^- is a $T_{[m-k+1]}$ if T^+ is a $T_{[k]}$, as can be seen in the case where $m = 2$ in figures 4a and 4b.

5. Objectives of the Discriminant Analyses and the Procedure Followed

5.1 The objectives

The analogy to a statistical test procedure pointed out in Section 3.3.2 makes clear that *exclusion* of a diagnostic category is the basis of discrimination on a statistical level. The greater the "power" of the discriminant region serving as the critical region, the greater is the certainty that detection follows from exclusion. A number of objectives for discriminant analysis follow from this observation:

1. The maximum attainable efficiency should always be determined. If the efficiency level attainable is high enough, one then has a discrimination criterion that can be applied symmetrically with reference to *both* categories.

2. If an adequate level of efficiency for this purpose cannot be attained, different kinds of exclusion with different degrees of certainty will be necessary, depending on the purpose.

In epidemiological investigations, for example, the exclusion of non-alcoholics, that is the detection of alcoholics, can be of primary importance. In this case a certain high specificity will be required and a T^+ sought with optimal sensitivity. In other situations the goal will be to exclude alcoholics, that is to detect non-alcoholics. In this case, discriminant regions T^+ with optimal specificity are sought when a certain high level of sensitivity is required. If the approach is based on simple discriminant regions that are symmetrical with respect to the variables, then the following parameters can be varied in the model:

- (i) The order k of the discriminant region T^+
- (ii) The set of variables to be investigated — with a view to using as few as possible
- (iii) The sector in which the multivariate decision point is to be sought

- (iv) The minimum efficiency, sensitivity and specificity required, since differences in the certainty of exclusion can be of practical importance, as is also the case with statistical tests.

A special role is played by the range of values permitted for the decision point d , and therefore special attention is given this topic in the next section.

5.2 Considerations in the selection of the multivariate decision limits

5.2.1 Optimization when the decision limits have been specified for each variable

In medicine, reference intervals (previously called normal ranges) and the associated location variables play an important, perhaps too important role. It may therefore be desirable, for instance, that for a certain set of variables one of the end points of each reference interval is specified in advance and for other variables the location parameter (or, more generally, some specific characteristic of the reference interval for each variable) is fixed.

In this case, the characteristics that can still be varied in this type of nonparametric discriminant analysis are the selection of the variable set and the order of the discriminant region.

5.2.2 Previously specified range for the multivariate decision point and the search for an optimum

It is, of course, usually much too restrictive to specify a single value as the decision limit for each variable, even if this value is of great importance for the clinician.

Such a value would have associated with it many uncertainties, at least those of a statistical and measurement nature. For each optimization problem an assessment must therefore be made of whether for the variables involved there are ranges in which it would be especially important to find an optimal multivariate decision point.

With the specification of a range for each decision limit and the selection of the discrimination criterion to be optimized (sensitivity, specificity, efficiency) the framework for the optimization search has been established. As a rule, the procedure will be iterative, i.e. it will contain an initial step that gives an indication of where the optimum could lie and additional steps that successively narrow down the optimum within the specified range.

5.2.3 The computer programme

Since there were no computer programmes available for problems of the type just described, we had to develop our own. In developing such programmes one is faced with the alternative of either automating *all* of the necessary steps or of designing the programme in such a way that one can look at each step and the results of that step. After running a few experimental programmes we decided to write a programme package that would permit two types of evaluation after certain factors had been specified:

First type of evaluation (see flow charts in figs. 5a and 5b):

First a subset is selected from the set of all available clinical chemical parameters. Then the desired order of the discriminant region and a "lattice" of decision points to be analysed are specified. A "lattice" of points in a space of a specified number of dimensions means any set of points that are equidistant from each other in each of the dimensions (whereby for different dimensions different distances are permitted – see the example below). In addition, either a lower limit for efficiency is specified or intervals are specified for both sensitivity and specificity. The computer programme then selects those points in the

previously defined lattice of points that attain or exceed the efficiency limit specified or whose sensitivity and specificity lie in the specified intervals. These points can be printed out with the coordinates, the sensitivity, the specificity and the efficiency.

Example:

Tables 5a and 5b show, in condensed form, the two-dimensional frequency distributions for the constituents γ -glutamyltransferase and mean corpuscular volume (MCV) for the non-alcoholics and alcoholics, respectively. To illustrate the differences between the two-dimensional distributions for the alcoholics and non-alcoholics simple discriminant regions of the order 2 have been selected – see figure 4a. In order to take advantage of the first type of evaluation just described, it was found useful to specify a kind of "lattice of lattices." An example follows:

Every combination of an a with a b represents a lattice of points in two dimensions. Table 6a shows part of a computer printout in which a lower limit of 155 is specified for the efficiency; a_1 and b_1 from table 5c have been combined. Table 7a shows, for the same combination, the printout when an interval of 93% to 100% has been specified for the sensitivity (speci-

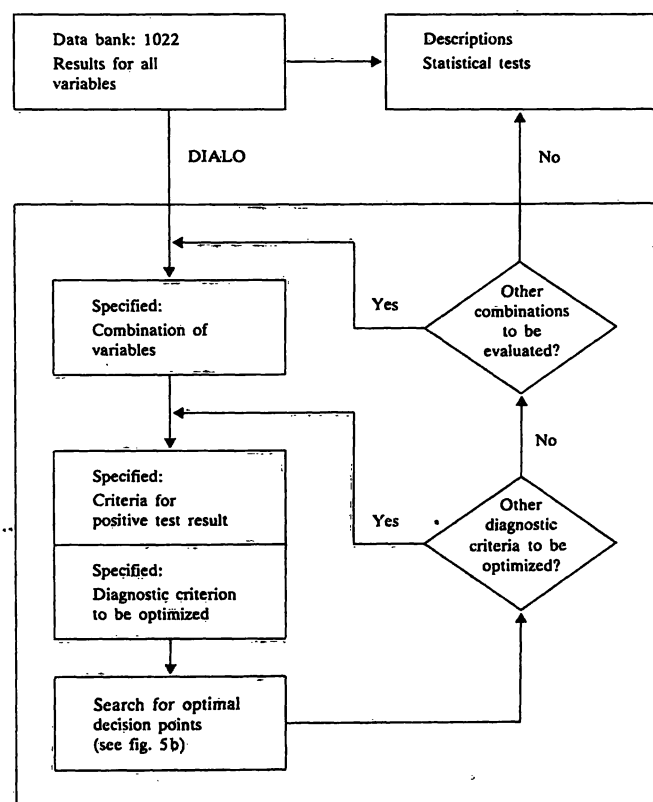


Fig. 5a. Flow chart for locating optimal multivariate decision limits.

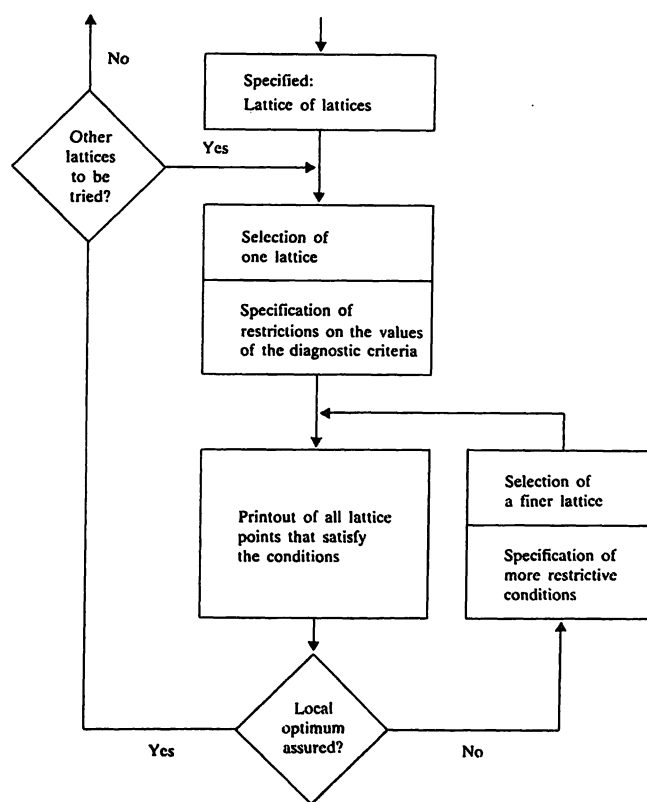


Fig. 5b. Detail of fig. 5a: Search for optimal decision limits in a lattice of lattices.

ficity: no restrictions). In the latter case the points associated with a sensitivity of exactly 95.06% are of special interest. Now in a second step a finer lattice is placed over the point or points with especially high specificity in order to find local maxima for specificity when sensitivity is 95.06%; in addition, the size of

the interval for the specificity has been reduced on the basis of the information gained in the first step. A similar procedure is carried out for the efficiency — see tables 6b and 7b. The first step always serves to provide a general overview, and the subsequent steps are used to determine the local maxima.

Tab. 5a. Two-dimensional frequency table of analytical results for the constituents γ -glutamyltransferase (GGT, U/l) and MCV (fl) (non-alcoholics).

GGT	MCV				
	72–80	81–89	90–98	99–107	≥ 108
0–19	1	6	17	8	1
20–39	2	11	5	7	2
40–59			4		
60–79		1	1		2
80–99	1				
100–199				1	
≥ 200					

Tab. 5b. Two-dimensional frequency table of analytical results for the constituents γ -glutamyltransferase (GGT, U/l) and MCV (fl) (alcoholics).

GGT	MCV				
	72–80	81–89	90–98	99–107	≥ 108
0–19			3	4	
20–39		1	6	13	1
40–59	1	1	7		2
60–79			6	2	
80–99		1		1	1
100–199			5	7	2
≥ 200			2	13	2

Tab. 5c. Two lattices each (a_1 and a_2 , b_1 and b_2) for the constituents γ -glutamyltransferase (GGT, U/l) and MCV (fl); by combining each a with each b a lattice of lattices is obtained.

GGT (U/l)		MCV (fl)	
a_1	$\left\{ \begin{array}{l} 12 \\ 15 \\ 18 \\ 21 \\ 24 \\ 27 \\ 30 \\ 33 \end{array} \right.$	b_1	$\left\{ \begin{array}{l} 80 \\ 82 \\ 84 \\ 86 \\ 88 \\ 90 \\ 92 \end{array} \right.$
a_2	$\left\{ \begin{array}{l} 35 \\ 40 \\ 45 \\ 50 \\ 55 \\ 60 \\ 65 \\ 70 \end{array} \right.$	b_2	$\left\{ \begin{array}{l} 94 \\ 96 \\ 98 \\ 100 \\ 102 \\ 104 \\ 106 \end{array} \right.$

Tab. 6a. Example in two dimensions of a first step in optimizing efficiency.

Specified:

At least two positive findings

Efficiency ≥ 155.00

Lattice: See tab. 5c, combination of a_1 and b_1

GGT = γ -glutamyltransferase

Pairs of decision limits GGT (U/l) MCV (fl)	Sensitivity (%)	Specificity (%)	Efficiency
18.00 88.00	91.36	64.29	155.64
21.00 88.00	86.42	70.00	156.42
18.00 90.00	90.12	68.57	158.69
21.00 90.00	85.19	72.86	158.04
24.00 90.00	79.01	77.14	156.16
27.00 90.00	76.54	80.00	156.54
18.00 92.00	82.72	74.29	157.00
21.00 92.00	77.78	78.57	156.35

Eight lattice points satisfy the above conditions.

Tab. 6b. Example in two dimensions of a second step in optimizing efficiency.

Specified:

At least two positive findings (see tab. 6a)

Efficiency ≥ 158.00

Lattice around the point GGT = 18 U/l, MCV = 90 fl

GGT = γ -glutamyltransferase

Pairs of decision limits GGT (U/l) MCV (fl)	Sensitivity (%)	Specificity (%)	Efficiency
18.00 90.00	90.12	68.57	158.69

One lattice point satisfies the above conditions.

Second type of evaluation:

After a single decision point has been selected the frequency distribution for all possible ways of exceeding the associated decision limits is given for the alcoholics and non-alcoholics. This includes: the number of cases in each sample in which neither coordinate exceeds the relevant decision limit, the number in which one coordinate exceeds the relevant decision limit, the number in which any two coordinates exceed the decision limits, and so on. Expressed in other terms, the frequency distribution of all combinations of "positive" and "negative" test results is determined for the alcoholics and non-alcoholics. Table 8 illustrates this kind of evaluation for

the case of two variables. For the point in table 6b with maximum efficiency, table 8 shows exactly how this efficiency is determined. There is also the possibility of having each value for each subject marked automatically on the computer printout as being positive or negative.

Tab. 7a. Example in two dimensions of a first step in optimizing specificity for a required level of sensitivity.
Specified: At least two positive findings
Sensitivity: 93%–100%
Specificity: 0%–100%
Lattice: See tab. 5c, combination of a_1 and b_1
GGT = γ -glutamyltransferase

Pairs of decision limits		Sensitivity (%)	Specificity (%)	Efficiency
GGT (U/l)	MCV (fl)			
12.00	80.00	97.53	18.57	116.10
15.00	80.00	95.06	37.14	132.20
18.00	80.00	93.83	47.14	140.97
12.00	82.00	97.53	20.00	117.53
15.00	82.00	95.06	37.14	132.20
18.00	82.00	93.83	47.14	140.97
12.00	84.00	97.53	22.86	120.39
15.00	84.00	95.06	40.00	135.06
18.00	84.00	93.83	50.00	143.83
12.00	86.00	95.06	34.29	129.35
12.00	88.00	95.06	40.00	135.06
12.00	90.00	93.83	44.29	138.11

Twelve lattice points satisfy the above conditions.

Tab. 7b. Example in two dimensions of a second step in optimizing specificity for a required level of sensitivity.
Specified:
At least two positive findings
Sensitivity: 93%–100%
Specificity: 40%–100%
Lattice around the points
GGT = 15 U/l, MCV = 84 fl
GGT = 12 U/l, MCV = 88 fl

Pairs of decision limits		Sensitivity (%)	Specificity (%)	Efficiency
GGT (U/l)	MCV (fl)			
16.00	83.00	95.06	40.00	135.06
17.00	83.00	93.83	44.29	138.11
15.00	84.00	95.06	40.00	135.06
16.00	84.00	95.06	41.43	136.49
17.00	84.00	93.83	45.71	139.54
14.00	85.00	93.83	41.43	135.26
13.00	86.00	95.06	42.86	137.92
14.00	86.00	93.83	47.14	140.97
13.00	87.00	95.06	42.86	137.92
14.00	87.00	93.83	47.14	140.97
12.00	88.00	95.06	40.00	135.06
13.00	88.00	95.06	47.14	142.20
14.00	88.00	93.83	51.43	145.26
12.00	89.00	95.06	41.43	136.49
13.00	89.00	95.06	48.57	143.63
14.00	89.00	93.83	52.86	146.68

Sixteen lattice points satisfy the above conditions.

Tab. 8. Frequency distribution for the various ways of exceeding the decision point GGT = 18 U/l, MCV = 90 fl from tab. 6b.
GGT = γ -glutamyltransferase

Possible combinations	A No.	A (%)	NA No.	NA (%)
No positive findings	0	0.0	5	7.1
Only GGT positive	4	4.9	18	25.7
Only MCV positive	4	4.9	25	35.7
GGT and MCV positive	73	90.1	22	31.4

5.2.4 Unlimited optimization

It cannot be expected that a range specified for the decision point d on the basis of clinical or traditional arguments will provide the best possible discrimination between two samples. If the range for d is left completely open, the search problem will only become more complicated, but it will still be solvable in principle with the programmes briefly described above. Computer-related limitations exist only with respect to external conditions such as available central processing unit (CPU) time. The only new methodological problem is to determine the absolute optimum from among the several local optima that usually exist. This requires a certain amount of care in the selection of the lattices for the decision point. As already mentioned, in the case of unlimited optimization it was found useful to work with a lattice of lattices.

We decided against more automatization of the analytical process, at least for the time being, in order not to risk losing any valuable information that might be obtained from intermediate steps. The "absolutely optimal" results sometimes appear to be rather peculiar in that for some constituents the decision limits are unexpectedly extreme. It must not be forgotten that there may be local optima in addition to the absolute optimum, and that they may be almost as good. On the other hand, there is usually a whole range around an optimum for which the discrimination criterion to be optimized has identical or almost identical values. To this point see also Comment 7.2.

5.3 The selection of the basic parameters

At the start of the study 30 clinical chemical parameters were available for evaluation (see l.c. (1)). It could be assumed from the outset that only a relatively small number of these parameters would make a significant contribution to good discrimination, and

in addition there was the express wish that good discrimination be achieved with as few variables as possible. On the other hand, it seemed desirable to know for which variables there are any differences at all between alcoholics and non-alcoholics. The first phase of the study therefore consisted of comparing the two distributions for every parameter, using a procedure that not only permitted determination of general differences between the distributions but also enabled a description of these differences. A very useful procedure for samples of sufficient size is the chi-square test, which has the additional advantage of being applicable even if the measurements have not been made using a continuous scale. The use of the same type of analysis for all of the data simplified the identification of those variables for which the differences between the distributions for the alcoholics and non-alcoholics are the greatest.

In the next phase the maximum efficiency attainable for each variable was determined. In both phases the same relatively small group of variables was found to be markedly better than all the rest. Although the two methods of comparison are related in a certain sense, this result was not predictable. We took it as reason enough to develop further analytical techniques on the basis of this group of variables, the constituents being:

GGT/GOT/GPT/GLDH/MCV/creatinine/urea-N.

GGT = γ -glutamyltransferase

GOT = aspartate aminotransferase

GPT = alanine aminotransferase

GLDH = glutamate dehydrogenase

MCV = mean corpuscular volume

In the course of subsequent analyses it became clear that the constituent glutamate dehydrogenase improved discrimination only slightly as compared with the other six, and therefore this parameter does not appear hereafter.

Of course the question must be raised of whether variables eliminated solely on the basis of one-dimensional observations might enable a much better discrimination in conjunction with other variables. This would surely be possible for certain types of distribution, but in our opinion not for the two discussed here. This is because, in the final analysis, one distribution is the result of the other in that certain variables change rather markedly in a certain direction.

6. Overview of the Results

Numerous typical results have been presented and discussed elsewhere (2). Since in this paper we are concerned more with methodological questions, we can limit ourselves here to a brief overview.

6.1 Dependence of the results on the constituent combination

If one begins with the variable that discriminates best and then successively adds the next best constituent, one can expect efficiency to improve gradually, just as is the case for sensitivity when a given level of specificity is required, and vice versa. Improvement will probably be quite marked at first, with subsequent steps bringing smaller and smaller amounts of change. This is confirmed in table 9 (see also l.c. (2), tab. 3–14), which shows the maximum efficiency achievable for combinations of up to six parameters, including the associated values for the linear discriminant analysis. A slight improvement can be obtained (but not with linear discriminant analysis) by using aspartate aminotransferase and the ratio aspartate aminotransferase/alanine aminotransferase instead of the pair aspartate aminotransferase and alanine aminotransferase.

Tab. 9. Maximum efficiency obtainable with simple discriminant regions and with linear discriminant analysis for several constituent combinations.

GGT = γ -glutamyltransferase

MCV = mean corpuscular volume

GOT = aspartate aminotransferase

GPT = alanine aminotransferase

Creat = creatinine

Constituent combination	Maximum efficiency		
	Simple discriminant region (Criterion: Minimum no. positive)		Linear discriminant analysis
GGT	150.5	(1)	—
GGT/MCV	158.7	(2)	142.6
GGT/GOT/MCV	165.2	(3)	144.8
GGT/GOT/MCV/urea-N	168.0	(4)	150.2
GGT/GOT/GPT/MCV/creat/urea-N	171.3	(5)	150.2
GGT/GOT/R*/MCV/creat/urea-N	173.6	(5)	146.0

* R = ratio GOT/GPT.

By including additional constituents the efficiency could be improved slightly (to about 176) with simple discriminant regions. The efficiency obtained with linear discriminant analysis was about 154.

6.2 Results for a combination of six parameters

In the previous section it was pointed out that the combination γ -glutamyltransferase, aspartate aminotransferase, alanine aminotransferase, mean corpuscular volume (MCV), creatinine and urea-N, which consists of constituents that are relatively easy to determine, yields results that are quite close to the best possible results obtainable for the given samples with simple discriminant regions. Table 10 shows a number of other results for this combination, namely the maximum values associated with various important levels of sensitivity and specificity (100%, 95%, 90% and 80%). It can be seen that the maximum achievable sensitivity for a given level of specificity is about the same as the maximum specificity for a given level of sensitivity.

Tab. 10. Maximum specificity obtainable for the combination GGT/GOT/GPT/MCV/creat/urea-N with different sensitivity requirements and vice versa.

GGT = γ -glutamyltransferase
MCV = mean corpuscular volume
GOT = aspartate aminotransferase
GPT = alanine aminotransferase
Creat = creatinine

Required sensitivity (%)	Maximum specificity (%)	Minimum no. positive
100.0	52.9	5
95.1	74.3	4
90.1	80.0	either 4 or 5
80.2	90.0	4
Required specificity (%)	Maximum sensitivity (%)	Minimum no. positive
100.0	55.6	5
95.7	69.1	5
90.0	80.2	4
80.0	90.1	either 4 or 5

6.3 Sensitivity and specificity charts

The programme package described in Section 5.2.3 would enable us to find, for a given combination of variables and for a given order of the discriminant region, all lattice points associated with a certain

level of sensitivity (specificity) and, for example, a minimum value for the specificity (sensitivity). Such "charts" would also be complete representations of the relationships between the two multidimensional empirical distributions that are to be differentiated from one another. They could be of value for a more refined kind of diagnosis involving more than decisions between two alternatives.

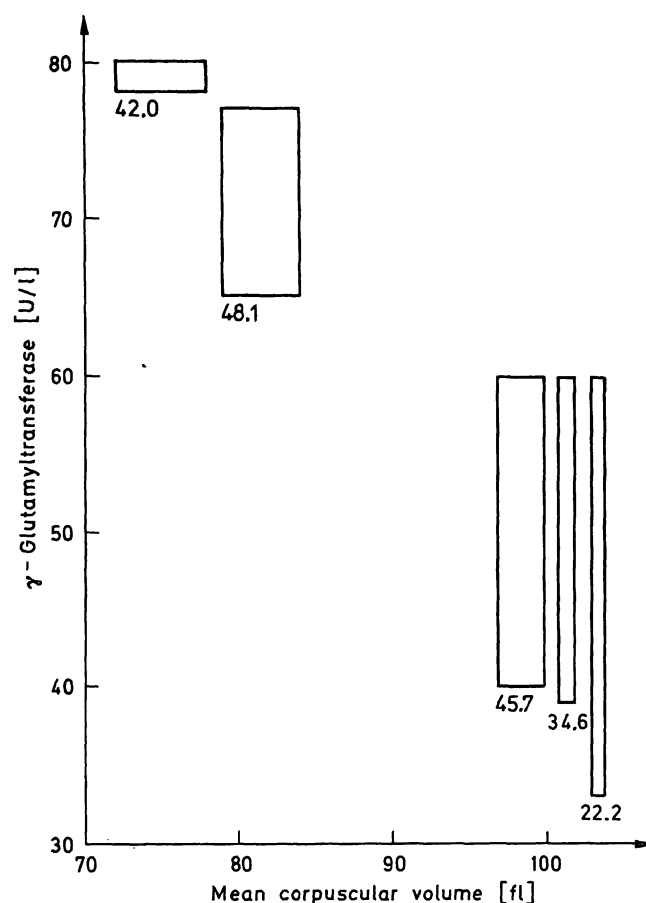


Fig. 6. Decision points for the combination γ -glutamyltransferase and mean corpuscular volume (MCV) with a specificity of 95.7% and a sensitivity of at least 20%.

Figure 6 shows a section of such a chart, with all points indicated for the combination γ -glutamyltransferase/mean corpuscular volume and the order 2 (i.e. at least two positive findings) of the discriminant region that are associated with a specificity of 95.71% and a sensitivity of at least 20%. The individual decision points have been combined into several rectangles, and the sensitivity associated with the lower left corner of each rectangle noted. The location with maximum sensitivity can thus also be identified. (See also Comment 7.4.)

7. Comments

This section includes comments on a few theoretical aspects of the methodology without the questions being addressed formally.

7.1 Simple discriminant regions of the k^{th} order and multidimensional quantiles

One of the most important aids in the general description of both one-dimensional probability distributions and one-dimensional empirical distributions is the quantile. The concept of a distribution function is based upon it. In the multidimensional case, the concept of a distribution function is usually based on the concept corresponding to a simple discriminant region of the order m or 1. And this is usually sufficient for deducing the common theorems about multidimensional probability distributions. The idea of introducing discriminant regions of any order suggests itself as a generalization. This means that the concept of the quantile can be defined in a number of different ways for more than two dimensions.

7.2 Description and estimation

The methodology presented here is a technique for describing the differences between empirical distributions, and the results are the values found for certain descriptive parameters. All of the values and each descriptive method must, however, be seen against the background of corresponding probability models, as already pointed out repeatedly (see, e.g., Section 2.1). This means that the descriptive parameters are to be regarded as estimators for corresponding parameters in probability models. The probability models are the mathematical representation of the random variability of the results when the same kinds of observations are made repeatedly in comparable situations.

All the analyses conducted in the present study involved a search for points with certain characteristics in a space with a specified number of dimensions, the points being based on the available results and the characteristics of the points being related to the values of the diagnostic criteria sensitivity, specificity and efficiency of simple discriminant regions of a given order. Or, in other words, what was sought was the number of points in either one (sensitivity, specificity) or two (efficiency) empirical distributions that were contained in a certain region. If such a region has been specified prior to the analysis, e.g. by characteristic values from reference intervals, we are concerned with a classic estimation problem. This is not the case for the optimization problems, however.

Probability statements about the properties of solutions to such optimization problems are therefore impossible, at least at present. However, all of the evaluations showed that the more dimensions involved, the hazier the local extremes, i.e. for the larger numbers of dimensions there is a certain amount of "play" around each local extreme in which the values of the optimized criterion are about the same.

It must also be kept in mind when considering the use of our findings elsewhere that these results are based on data from a certain constellation of institutions (see Section 2.2). For a different constellation or for individual institutions, all the results might be somewhat better or worse (see Section 2.2.3), even if they were obtained under comparable conditions.

7.3 Visualizing and interpreting statistical results

Geometric structures in more than three dimensions cannot be visualized. A linear discriminant function in two or three dimensions can be visualized easily, whereas one in six dimensions "lives" from consideration of analogous situations or from the trick of projection onto two-dimensional subspaces. However, neither of these techniques conveys a direct, complete picture.

A statistical result, i.e. the final result of a statistical evaluation, should have a relationship to something "real," i.e. to something that can be approached or made accessible through observation and has some concrete meaning. The calculation of a mean of 2.51 on a four-point ordinal scale is, for example, uninterpretable, since there is no way of presenting for examination an object that would correspond to such a value on the given scale. In using a linear (or quadratic or any other) discriminant function the individual values for each patient are transformed into a value of this function. This value leads to the assignment of a patient to one of the given categories, but knowledge of this value no longer implies knowledge of what it is based on, and thus the value is barely interpretable.

But in our view the properties of clarity and interpretability just outlined are of great importance, especially for the clinician. We therefore want to emphasize that the discrimination method used provides results that can be both visualized and interpreted. Now it is true that a simple discriminant region of the 4th order in six dimensions cannot be visualized as a geometric structure except via analogy, but this structure corresponds to a series of logi-

cal connectives that can be stated explicitly. The corresponding discrimination rule is that a set of values is "positive" when at least four of the six values exceed the associated decision limits. Such a situation can be visualized easily. But since the methodology also shows which limits have been exceeded, the particular combination of positive and negative results is known and can then be considered an indicator of, for example, possible organ damage. This is precisely what is meant by interpretability.

7.4 The flexibility of the methodology and the exclusion principle

Common procedures for discriminant analysis are usually programmed to maximize the percentage of correct assignments (see Section 4.1). It is impossible with currently available programmes to solve the other optimization problems mentioned in Section 5.1. With the approach reported here, however, this is relatively easy, the reason being, as mentioned in Section 7.1, that in the final analysis the method involves the determination of multidimensional quantiles.

Such flexibility is, however, an important property of discrimination procedures, for, as mentioned in Section 3.3.2, the importance of discrimination criteria lies in the analogy with statistical test criteria, i.e. the exclusion principle is the key element in their use. But a statistical test must be applicable for different test levels depending on the problem at hand, and, analogously, different situations will require the setting of different levels of sensitivity and/or specificity.

Moreover, the diagnostic significance of a result must be regarded as something like a continuous variable, just as is the case for the statistical significance of a test, for example a test for comparing two means. It is true that in the usual form of a statistical test, the "significance" of a difference between two means is categorized as given or not given on the basis of a specified test level. But it is also a truism,

though unexpressed, that a difference between means corresponding to an 0.5% significance level (i.e. a test level of 0.5%), has a "greater" significance than a difference with a test level just under 5%. Therefore analogous differences in diagnostic significance can certainly be assumed for different locations of a point within the range of positive or negative values. This implies that it should be possible in principle to produce a chart of critical values for sensitivity and/or specificity that is analogous to a table of critical values for a statistical test.

7.5 Multidimensional decision limits for more than two categories

Our methodology in its present form has the great disadvantage that it is designed to discriminate between two categories only. This does not mean that there would be insurmountable difficulties in using it to discriminate among more than two categories but rather that the method would have to be adapted for this purpose; the treatment of such problems would require several multidimensional decision limits. Such an approach appears to us to be feasible in principle, and we can also envisage several different ways of tackling the problem.

We will report more on this problem as soon as the necessary adaptation of concepts and development of computer programmes have been completed.

Acknowledgments

Special thanks to Ms. *Ingrid Bickert* and Ms. *Eva Müller* (Department of Biostatistics) for the great care with which they analysed the many "lattices of lattice points," a process requiring great persistence, especially during the early phases of method development.

Special thanks also to Ms. *Ursula Grau* (Department of Biostatistics) for the patience and care with which she prepared numerous versions of the manuscript as we attempted to present our methodology in an understandable form.

This methodology could not have been developed without the continued and active collaboration of the Department of Clinical Chemistry, the Psychiatric Outpatient Department and the Department of Biostatistics at the Max-Planck-Institut für Psychiatrie.

References

1. Stamm, D., Hansert, E. & Feuerlein, W. (1984) Excessive consumption of alcohol in men as a biological influence factor in clinical laboratory investigations. *J. Clin. Chem. Clin. Biochem.* 22, 65–77.
2. Stamm, D., Hansert, E. & Feuerlein, W. (1984) Detection and exclusion of alcoholism in men on the basis of clinical laboratory findings. *J. Clin. Chem. Clin. Biochem.* 22, 79–96.
3. Solberg, H. E. (1975) Discriminant analysis in clinical chemistry (editorial). *Scand. J. Clin. Lab. Invest.* 35, 705–712.
4. Suppes, P. & Zinnes, J. L. (1963) Basic measurement theory. In: *Handbook of Mathematical Psychology*, Vol. 1 (Luce, R. D., Bush, R. R. & Galanter, E., eds.). John Wiley & Sons, New York, pp. 1–76.
5. Stoermer, H. (1964) Ein Test zum Erkennen von Normalverteilungen. *Z. Wahrscheinlichkeitstheorie* 2, 420–428.
6. Büttner, H. (1982) Anwendung entscheidungstheoretischer Methoden. In: *Strategien für den Einsatz klinisch-chemischer Untersuchungen* (Lang, H., Rick, W. & Büttner, H., eds.). Deutsche Gesellschaft für Klinische Chemie, Merck-Symposium 1981. Springer-Verlag, Berlin, Heidelberg, New York, Discussion pp. 79–87.

7. Keller, H. & Gessner, U. (1982) The Risk Quotient: A Tool for Interpretative Reporting. Vortrag auf dem 5. Colloque Internationale de Biologie Prospective, 4.-9. Okt. 1982, Pont-à-Mousson. Unpublished manuscript.
8. Kollegium Biomathematik NW (1975) Biomathematik für Mediziner. Springer-Verlag, Berlin, Heidelberg, New York.
9. Armitage, P. (1971) Statistical Methods in Medical Research. Blackwell Scientific Publications, Oxford, Edinburgh.
10. Galen, R. S. & Gambino, S. R. (1975) Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, London, Sidney, Toronto.
11. Büttner, H. (1980) Grundlagen der Bewertung klinisch-chemischer Befunde. In: Validität klinisch-chemischer Befunde (Lang, H., Rick, W. & Büttner, H., eds.). Deutsche Gesellschaft für Klinische Chemie, Merck-Symposium 1979. Springer-Verlag, Berlin, Heidelberg, New York, pp. 58-72.
12. Anderson, T. W. (1958) An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York, London, Sidney.
13. Birkenfeld, W. (1978) One step towards a nonparametric discriminant analysis. In: COMPSTAT 1978, Proceedings in Computational Statistics, 3rd Symposium held in Leiden 1978 (Corsten, L. C. A. & Hermans, J., eds.). Physica-Verlag, Vienna, pp. 162-169.

Dr. E. Hansert
Max-Planck-Institut für Psychiatrie
Kraepelinstraße 10
D-8000 München 40